ELSEVIER

Journal of Marine Systems 16 (1998) 51–68

JOURNAL OF MARINE SYSTEMS

Data assimilation and a pelagic ecosystem model: parameterization using time series observations

Y.H. Spitz [a,*], J.R. Moisan [b], M.R. Abbott [a], J.G. Richman [a]

[a] *College of Oceanic and Atmospheric Sciences Oregon State University, Corvallis, OR, USA*
[b] *Physical Oceanography Research Division, Scripps Institution of Oceanography, La Jolla, CA, USA*

Received 24 October 1996; accepted 24 April 1997

Abstract

Variational adjoint assimilation of time series observations is used to estimate the optimal parameters of a nitrogen-budget, upper ocean, mixed-layer ecosystem model. Observations collected at the Bermuda Atlantic Time-Series Study (BATS) site are taken as an example of a time series. A twin experiment using simulated data of the same type and frequency as the BATS observations first demonstrates the adequacy of the observations to estimate the model parameters and model the ecosystem annual cycle. This experiment further shows that some of the model parameters cannot be estimated independently. This conclusion leads to a simplification of the model and a redefinition of its parameters. Based upon the success of the twin experiment to estimate all model parameters, an attempt to assimilate actual observations from BATS was undertaken. The assimilation of real data leads to the conclusion that, even though the frequency and type of observations is adequate to estimate the model parameters, the considered model is not appropriate for the annual cycle of the BATS ecosystem.

Keywords: models; BATS; time series analysis; pelagic environment; ecosystems

1. Introduction

Physical–biological models of various levels of sophistication have recently been developed for different regions of the ocean. The biological model was configured first as a compartmental ecosystem model in an upper ocean mixed-layer (e.g. Fasham et al., 1990). This class of ecosystem models has then been coupled to one-dimensional physical models (McGillicuddy et al., 1995a; Prunet et al., 1996a,b; Oguz et al., 1996), and embedded into three-dimensional circulation models (Fasham et al., 1993; Sarmiento et al., 1993; McGillicuddy et al., 1995b; Moisan et al., 1996). The main difficulty with these models is obtaining an estimate of the parameters. These parameters, such as zooplankton grazing, specific growth and mortality rates, are often poorly measured or poorly known. Typically, these parameters are individually and non-systematically adjusted until the model results 'fit' the observations. Such a subjective technique for 'fitting' the model parameters to the observations quickly becomes an arduous task, even with an ecosystem model that has few coefficients. As Fasham et al. (1990) pointed out in

* Corresponding author. Tel.: +1 (541) 737-3227; Fax: +1 (541) 737-2064; E-mail: spitz@oce.orst.edu

0924-7963/98/$19.00
PII S0924-7963(97)00099-7

their modeling study, more than 200 individual model runs were required in order to obtain a set of model parameters that led to results which compared well to the observations.

A systematic and non-subjective technique of adjusting the model parameters consists of using observations in conjunction with a data assimilation technique such as the simulated annealing (Armstrong et al., 1995; Matear, 1995) or variational adjoint method (Lawson et al., 1995, 1996). Data assimilation has been widely implemented in meteorological and ocean circulation studies (Ghil and Malanotte-Rizzoli, 1991; Malanotte-Rizzoli, 1996). More recently, data assimilation has been extended to coupled physical–biological models. Matear (1995) used a simulated annealing technique with data from Ocean Weather Station (OWS) Papa to estimate the parameters in three different biological models. Results from this work demonstrated that non-independent parameters within a model cannot be separately determined, but can be determined as combinations of each other, such as sums, products or fractions. Prunet et al. (1996a) applied a simple variational data assimilation method to calibrate a biological model for OWS Papa. They found that only linear combinations of the ecosystem model parameters can be adjusted, their values depending upon their a priori values. Fasham and Evans (1995) used a non-linear optimization technique and a seven-component ecosystem model in an upper ocean mixed-layer to fit the observations at the US Joint Global Ocean Flux Study (JGOFS) station at 47°N 20°W. They could not reach a solution which simultaneously gave a good fit to the primary production and the zooplankton observations and concluded that the mixed-layer nitrogen flows were not adequately modeled. Data assimilation using a variational adjoint method has also been applied to a five-component, time-dependent, ecosystem model to estimate population growth and mortality rates, amplitudes of forcing events and initial conditions (Lawson et al., 1996). Using identical twin experiments and sampling strategies corresponding to those of the US JGOFS experiments at the Bermuda Atlantic Time-Series (BATS) and the Hawaii Ocean Time series (HOT) stations, Lawson et al. (1996) investigated the effect of changes in data type and distribution on the ability to recover model parameters.

The main focus of this study is to assess the feasibility of using a data assimilation technique with sparse time series observations such as from BATS to estimate the poorly known parameters for the annual cycle of a nitrogen budget model in the upper ocean mixed-layer. In order to achieve this, two groups of data assimilation experiments were carried out. First, twin experiments using model-generated observations were run to determine if the frequency at which data are collected at BATS is sufficient to estimate all the model parameters. Second, data assimilation of the BATS data (1988–1993) was attempted. The pelagic ecosystem model for this study is based upon the Fasham et al. (1990) model that was previously subjectively calibrated using data from Hydrostation S.

The data assimilation technique, ecosystem model, and definition of the cost function and data availability are presented in Section 2. The results of the twin experiment and the assimilation of the BATS observations are presented in Sections 3 and 4, respectively. A summary of the results is given in Section 5.

2. Determination of the model parameters

2.1. Methodology

During the past decade, the variational adjoint method has been largely used in meteorology and oceanography to estimate initial and boundary conditions. It has since been used to estimate parameters in circulation models (Smedstad and O'Brien, 1991; Spitz, 1995), and an ecosystem model (Lawson et al., 1995, 1996). Two advantages of this technique are that it can be applied to both linear and non-linear models, and it can be implemented in a straightforward manner. Since the technique has been largely discussed in the literature, we will limit ourselves to a brief overview of the technique.

The variational adjoint method determines an optimal solution by minimizing an objective function, the cost function, which measures the difference between the model solution and the available observations. Most minimization algorithms require the computation of the gradient of the cost function with respect to the control variables, e.g. model parame-

ters. The data assimilative model consists of three components: the forward ecosystem model, the backward model or adjoint model, and an optimization procedure (Fig. 1). The three components of the assimilative model are used in an iterative procedure which leads to the determination of the control variables giving the best fit to the data and can be described as follows. The direct model is run with an initial guess of the control variables. The model output and data are then used to compute the value of the cost function. Thereafter, the adjoint of the model, run backward in time, gives the gradient of the cost function with respect to the control variables, which is then used in the optimization procedure to compute the search direction towards the minimum and the optimal step size in that direction. New values of the control variables are then estimated, and the model is rerun. This procedure is applied until a preset convergence criterion is satisfied, e.g. $J \leq \epsilon_1$ and/or $\|\nabla J\| \leq \epsilon_2$, where ϵ denotes a small value, J is the cost function and $\|\nabla J\|$ is the norm of the gradient of the cost function with respect to the control variables. In practice, to assure that the cost function has reached a global minimum and not a local minimum, the procedure is repeated with different first guesses of the control variables. The global minimum is reached if all the runs converge towards the same value of the cost function and the same control variables.

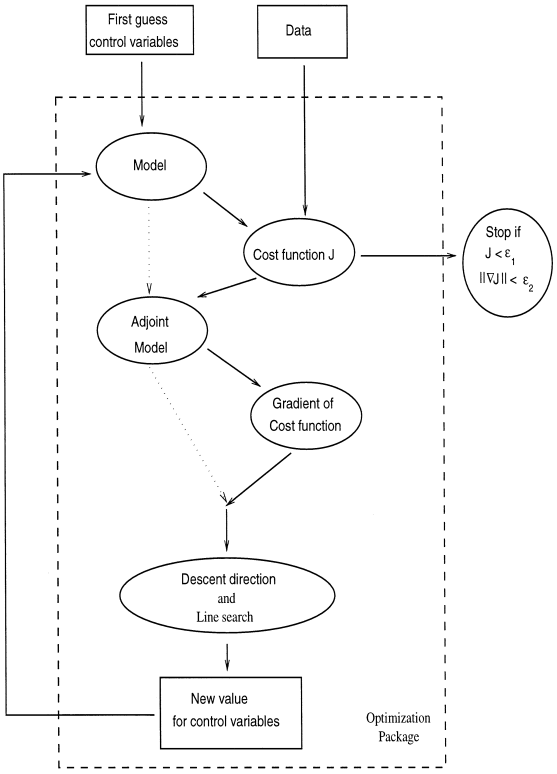

Fig. 1. Schematic of the steps involved in the data assimilation scheme. The solid lines indicate the main path taken during the procedure.

While the adjoint method is a powerful tool for obtaining the gradient of the cost function with respect to the parameters of the model, the most difficult aspect of this technique is the development of the adjoint model code. One approach consists of deriving the continuous adjoint equations followed by their discretization (Smedstad and O'Brien, 1991). Another approach is to derive the adjoint model code directly from the model code. In this case, the adjoint model code can either be built from the tangent linear model code (Spitz, 1995) or it can be constructed based upon the use of Lagrange multipliers. The second method was adopted in this study and a full description and an application to a simple prey–predator model can be found in Lawson et al. (1995). This technique provides a straightforward way of writing code and avoids the inconsistency that can occur from the discretization of the adjoint continuous equations. The optimization procedure uses the subroutine N1QN3 from Gilbert and Lemaréchal (1989), which is based upon a limited memory quasi-Newton method.

2.2. Ecosystem model

The ecosystem model in our study is the nitrogen budget model of the upper ocean pelagic ecosystem developed by Fasham et al. (1990). This model simulates the seasonal cycle of the ecosystem within the upper ocean mixed-layer. A system of seven coupled ordinary differential equations controls the time rate of change of the ecosystem constituents which include: nitrate (NO_3), ammonium (NH_4), dissolved organic nitrogen (DON), detritus (D), bacteria (B), phytoplankton (P) and zooplankton (Z). All model constituents are expressed in terms of mmol N m^{-3}. The model numerical scheme is a fourth-order Runge-Kutta with a 2-h time step.

The model is forced by a prescribed seasonally varying mixed-layer depth (MLD). Within the mixed-layer, the ecosystem is considered homogenous. The change in MLD over time:

$$\frac{\mathrm{dMLD}}{\mathrm{d}t} = h(t) \tag{1}$$

is calculated from the monthly mean mixed-layer depth at Bermuda using the Levitus (1982) data set. This change is further used to determine the rate of entrainment of water from greater depths into the mixed-layer:

$$h^{+}(t) = \max(h(t),0) \tag{2}$$

Finally, diffusive mixing between the mixed-layer and the deep ocean is parameterized by the inclusion of a vertical flux rate, m.

The time rate of change in the phytoplankton nitrogen pool (P) is:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = (1-\gamma_1)\sigma P - G_1 - \mu_1 P - \frac{m + h^{+}(t)}{\mathrm{MLD}} P \tag{3}$$

where σ is the specific light and nutrient limited growth rate, γ_1 is the phytoplankton exudation rate, G_1 is the zooplankton grazing rate, and μ_1 is the specific mortality rate. The last term in the equation represents the volumetric dilution of the phytoplankton concentration caused by the deepening of the mixed-layer, $h^{+}(t)$, and diffusive mixing, m.

The specific light and nutrient limited growth rate is defined as:

$$\sigma = J_I Q \tag{4}$$

where J_I and Q are the light- and nutrient-limited growth terms, respectively. The nutrient-limited growth term is given by:

$$Q = Q_1 + Q_2 \tag{5}$$

where Q_1 is the nitrate uptake:

$$Q_1 = \frac{\mathrm{NO_3} \mathrm{e}^{-\Psi \mathrm{NH_4}}}{K_1 + \mathrm{NO_3}} \tag{6}$$

and Q_2 is the ammonium uptake:

$$Q_2 = \frac{\mathrm{NH_4}}{K_2 + \mathrm{NH_4}} \tag{7}$$

and where Ψ is a parameterization of the strength of ammonium inhibition of nitrate uptake observed by Walsh and Dugdale (1972), and K_1 and K_2 are the half-saturation constants for nitrate and ammonium uptake, respectively.

The light-dependent growth rate takes into account the diel light cycle and is defined as:

$$J_I = \frac{1}{\mathrm{MLD}} \int_0^{\mathrm{MLD}} \frac{V_\mathrm{p}\, \alpha I_0 \mathrm{e}^{-(k_\mathrm{w} + k_\mathrm{c} P)z}}{\sqrt{\left(V_\mathrm{p}^2 + \alpha^2 \left(I_0 \mathrm{e}^{-(k_\mathrm{w} + k_\mathrm{c} P)z}\right)^2\right)}} \mathrm{d}z \tag{8}$$

where V_p is the maximum growth rate and α is the initial slope of the primary production versus light curve. The amount of incident photosynthetically available radiance (PAR) at the surface of the ocean, I_0, is attenuated by the mean PAR absorption coefficient for pure water, k_w, and the chlorophyll-specific mean PAR absorption coefficient, k_c.

The time rate of change of the zooplankton nitrogen pool (Z) is:

$$\frac{\mathrm{dZ}}{\mathrm{d}t} = \beta_1 G_1 + \beta_2 G_2 + \beta_3 G_3 - \mu_2 Z - \mu_5 Z - \frac{h(t)}{\mathrm{MLD}} Z \tag{9}$$

where β_1, β_2, β_3 are the assimilation efficiencies and G_1, G_2, G_3 are the grazing rates for the phytoplankton, bacteria, and detritus, respectively, and μ_2 and μ_5 are the specific excretion and specific mortality rates for zooplankton, respectively. The forms for the zooplankton grazing rates are given as:

$$G_1 = \frac{g Z p_1 P^2}{K_3(p_1 P + p_2 B + p_3 D) + p_1 P^2 + p_2 B^2 + p_3 D^2} \tag{10}$$

$$G_2 = \frac{g Z p_2 B^2}{K_3(p_1 P + p_2 B + p_3 D) + p_1 P^2 + p_2 B^2 + p_3 D^2} \tag{11}$$

$$G_3 = \frac{g Z p_3 D^2}{K_3(p_1 P + p_2 B + p_3 D) + p_1 P^2 + p_2 B^2 + p_3 D^2} \tag{12}$$

where g is the maximum grazing rate, K_3 is the half-saturation coefficient for grazing and p_1, p_2

and p_3 denote the zooplankton food preference for phytoplankton, bacteria, and detritus, respectively.

The time rate of change in the nitrogen pool of active bacteria (B) is:

$$\frac{\mathrm{d}B}{\mathrm{d}t} = U_1 + U_2 - G_2 - \mu_3 B - \frac{m + h^+(t)}{\mathrm{MLD}} B \quad (13)$$

where U_1 is the uptake of DON:

$$U_1 = \frac{V_b B \mathrm{DON}}{K_4 + S + \mathrm{DON}} \quad (14)$$

U_2 is the uptake of NH_4:

$$U_2 = \frac{V_\mathrm{b} B S}{K_4 + S + \mathrm{DON}} \quad (15)$$

μ_3 is the bacteria-specific excretion rate, V_b is the maximum bacterial uptake rate, and K_4 is the half-saturation coefficient for uptake. The partitioning between bacterial uptake of DON and NH_4 is controlled by the amount of total bacterial nitrogenous substrate:

$$S = \min(\mathrm{NH}_4, \eta \mathrm{DON}) \quad (16)$$

where η is the ratio of bacterial uptake of NH_4 to bacterial uptake of DON, as derived in Fasham et al. (1990).

The change in the detritus nitrogen pool (D) over time is:

$$\frac{\mathrm{d}D}{\mathrm{d}t} = (1-\beta_1)G_1 + (1-\beta_2)G_2 - \beta_3 G_3 - \mu_4 D + \mu_1 P - \frac{m + h^+(t) + V}{\mathrm{MLD}} D \quad (17)$$

where the first three terms parameterize the egestion of food material ingested by the zooplankton, μ_4 is the rate of detritus breakdown by bacterial enzymes and non-bacterial mediated processes, and V is the sinking rate of the detritus material.

The change in the nitrate pool (NO_3) over time:

$$\frac{\mathrm{dNO}_3}{\mathrm{d}t} = -J_I Q_1 P + \frac{m + h^+(t)}{\mathrm{MLD}} (N_\mathrm{bottom} - \mathrm{NO}_3) \quad (18)$$

where the flux of nitrate into the mixed-layer is controlled by the parameterized diffusive mixing flux, m, the rate of entrainment of deep ocean water, $h^+(t)$, and the difference between the nitrate concentration within the mixed-layer and that observed at just below the mixed-layer, N_bottom.

The time rate of change of the ammonium (NH_4):

$$\frac{\mathrm{dNH}_4}{\mathrm{d}t} = -J_I Q_2 P - U_2 + \mu_3 B + (\epsilon\mu_2 + (1-\Omega)\mu_5) Z - \frac{m + h^+(t)}{\mathrm{MLD}} \mathrm{NH}_4 \quad (19)$$

where Ω is the fraction of the remineralized zooplankton grazing mortality, and ϵ is the ammonium fraction of the zooplankton excretion.

The change in the dissolved organic nitrogen (DON) over time:

$$\frac{\mathrm{dDON}}{\mathrm{d}t} = \gamma_1 J_I Q P + \mu_4 D + (1-\epsilon)\mu_2 Z - U_1 - \frac{m + h^+(t)}{\mathrm{MLD}} \mathrm{DON} \quad (20)$$

where all the terms have been previously defined.

2.3. Cost function and data availability

The assimilation is performed by minimizing a cost function which measures the distance between observations and model equivalents to the observations. In this study, the cost function is defined in a least square manner as:

$$J = \frac{1}{2} \sum_{i,n} (d_{i,n} - a_{i,n})^T \boldsymbol{W}_i (d_{i,n} - a_{i,n}) \quad (21)$$

where d and a are the data and model equivalents to the data, respectively, i refers to the data types and n refers to the observation time. The model equivalent to the data were linearly interpolated to the time of the collection and converted to the same units as the observations. The conversion factors are described further in this section. The weighting matrices $\boldsymbol{W}_i$ are theoretically the inverse of the observation error covariance matrices. By assuming that errors in the data are uncorrelated and have equal variance, the weight matrices can be rewritten as:

$$\boldsymbol{W}_i = w_i \boldsymbol{I} \quad (22)$$

where w_i is a positive scalar. In practice, w_i takes into account the relative magnitude of the various data types and the quality of the data sets. In this study, w_i accounts for the differences in the relative magnitude observed in the time-average of each data type and is defined as:

$$w_i = \frac{\max\left(\bar{d}_j\right)}{\bar{d}_i} \tag{23}$$

where $\bar{d}_i$ is the time-average of the observation i and $\max(\bar{d}_j)$ is the maximum of the time-average of the assimilated observations (Lawson et al., 1996).

Data with a large uncertainty in the measurement and the conversion factor into nitrogen units are, however, assigned with a weight reduced by a factor 10 with respect to the value given by Eq. (23). Those data are described later in this section.

The data used in the assimilation process are equivalent to the observations taken at BATS and consist of nitrate, chlorophyll-*a* and particulate organic nitrogen concentrations, bacteria cell counts, phytoplankton primary production rates, and bacterial production rates. Sampling at this site began in October 1988 and occurs at about monthly intervals. A full description of the BATS observations can be found in Michaels and Knap (1996) and a list of the data collected is shown in Table 1.

The data from BATS are depth-dependent observations collected from the surface down to greater than 250 m in depth. However, the ecosystem model simulates the ecosystem concentrations and processes as homogeneous within the mixed-layer. Each BATS data profile was vertically integrated from the surface to the seasonally varying mixed-layer depth (MLD) to make the BATS data compatible with the ecosystem model solutions.

The actual data used in the assimilation procedure, such as bacterial cell counts or chlorophyll-*a* concentration, are not prognostically determined in the Fasham et al. (1990) model. In order to make the model output compatible with the BATS observations a number of assumptions, data conversions and diagnostic equations were needed.

The observed phytoplankton chlorophyll-*a* data were assumed to be a good representation of the amount of phytoplankton biomass. The model phytoplankton nitrogen concentrations were then converted from the model units (mmol N m^{-3}) to chlorophyll concentration (mg chlorophyll-*a* m^{-3}) using an assumed constant chlorophyll-*a* to nitrogen ratio of 1.59 g chlorophyll-*a* (mol N)$^{-1}$, which was calculated using a constant carbon to nitrogen (C:N; mol:mol) ratio of 6.625 and a carbon to chlorophyll-*a* (g:g) ratio of 50.

The bacteria data derived from BATS were recorded as cell counts per unit mass and the bacte-

Table 1
Type of data collected at US JGOFS Bermuda Atlantic Time-Series Study (BATS) station between 1988 and 1993

Value measured	Measurement method	Units
Salinity	CTD, salinometer	n.d.
Temperature	CTD	°C
Density	Calculated	kg m^{-3}
Dissolved oxygen	Winkler titration	μmol kg^{-1}
Total CO_2	Coulometric technique	μmol kg^{-1}
Nitrate	Autoanalyzer	μmol kg^{-1}
Nitrite	Autoanalyzer	μmol kg^{-1}
Phosphate	Autoanalyzer	μmol kg^{-1}
Silicate	Autoanalyzer	μmol kg^{-1}
Chlorophyll-*a*	Fluorometer and HPLC	μg kg^{-1}
Phaeopigments	Fluorometer and HPLC	μg kg^{-1}
Bacterioplankton	Enumeration *w*/DAPI	cell kg^{-1}
PON	Elemental analyzer	μg kg^{-1}
POC	Elemental analyzer	μg kg^{-1}
Primary production	^{14}C uptake method	mg C m^{-3} d^{-1}
Bacterial production	^{3}H-thymidine uptake method	pmol l^{-1} h^{-1}

rial production rates were measured in terms of [^{3}H-methyl]thymidine (^{3}H-TdR) uptake rates. Conversions were needed in order to make the simulated data comparable to the observed data. However, Carlson et al. (1996) pointed out that there is no universally accepted conversion factor for conversion of bacterial abundance to carbon biomass or ^{3}H-TdR uptake to bacterial carbon production. Carlson et al. (1996) also pointed out that bacteria biomass estimates are approximately 10% overestimated due to the inclusion of prochlorophyte cells in the total bacteria cell counts, while Sieracki et al. (1995) found that 10–25% of the counted cells are prochlorophytes and not heterotrophic bacteria. Even with these difficulties, we assumed a constant carbon to nitrogen (C:N) ratio of 4 and a bacteria cell biomass of 20 fg C cell^{-1} to calculate a conversion factor of 0.048 mmol N m^{-3} (10^8 cells kg^{-1})$^{-1}$. Considering the uncertainties in conversion factor and measurement, the bacteria cell counts were given a reduced weight, i.e. one order of magnitude smaller, in the cost function and bacterial production was not assimilated.

In order to equate the particulate organic nitrogen (PON) measurements from BATS to the model prognostic variables, a diagnostic equation for PON was written as:

$$\mathrm{PON} = P + Z + B + \mathrm{D} \tag{24}$$

where P, Z, B and D are in terms of mmol N m^{-3}. The model phytoplankton net primary production rate which is compared to the BATS phytoplankton primary production rate is defined as:

$$\mathrm{PP} = (1 - \gamma_1)\,\sigma P \tag{25}$$

where σ is the specific light and nutrient limited growth rate and γ_1 is the phytoplankton exudation rate. A C:N (mol:mol) ratio of 6.625 was also used to convert the simulated phytoplankton nitrogen uptake estimates to carbon uptake estimates. The con-

Table 2
Fasham et al. (1990) model parameters

Parameter	Symbol	Value	Dimension
Light attenuation due to water	k_w	0.04	m^{-1}
Cross-thermocline mixing rate	m	0.1	m d^{-1}
Phytoplankton maximum growth rate	V_p	2.9	d^{-1}
Initial slope of the P/I curve	α	0.025	(w m^{-2})$^{-1}$d^{-1}
Half-saturation for phytoplankton NO_3 uptake	K_1	0.5	mmol N m^{-3}
Half-saturation for phytoplankton NH_4 uptake	K_2	0.5	mmol N m^{-3}
Phytoplankton specific mortality rate	μ_1	0.09	d^{-1}
Light attenuation by chlorophyll-*a*	k_c	0.03	m^2 (mmol N)$^{-1}$
Phytoplankton exudation rate	γ_1	5%	none
NH_4 inhibition parameter	Ψ	1.5	(mmol N)$^{-1}$
Zooplankton maximum growth rate	g	1.0	d^{-1}
Zooplankton assimilation efficiencies	β_1, β_2, β_3	75%	none
Zooplankton specific excretion rate	μ_2	0.1	d^{-1}
Zooplankton specific mortality rate	μ_5	0.05	d^{-1}
Zooplankton half-saturation for ingestion	K_3	1.0	mmol N m^{-3}
Detritus fraction for zooplankton mortality	Ω	33%	none
NH_4 fraction of zooplankton excretion	ϵ	75%	none
Bacteria maximum growth rate	V_b	2.0	d^{-1}
Bacteria-specific excretion rate	μ_3	0.05	d^{-1}
Detrital breakdown rate	μ_4	0.05	d^{-1}
Bacteria half-saturation rate for uptake	K_4	0.5	mmol N m^{-3}
NH_4/DON uptake ratio	η	0.6	none
Detritus sinking rate	V	1.0	m d^{-1}
Zooplankton food preference for phytoplankton, bacteria, detritus	p_1, p_2, p_3	1.0, 1.0,1.0	none
Nitrate concentration below z = MDL	N_{bottom}	2.0	mmol N m^{-3}

version assumes that the phytoplankton were always in a state of balanced growth.

3. Twin experiment

In order to systematically test the ability of the assimilation technique to recover parameters, numerical experiments with simulated data (twin experiments) were carried out. The first issue to be addressed with the twin experiments is the feasibility of using sparse observations such as obtained by BATS to estimate the various model parameters (see Table 2) and model the ecosystem annual cycle. Matear (1995) argued that more than half of the Fasham et al. (1990) model parameters are highly

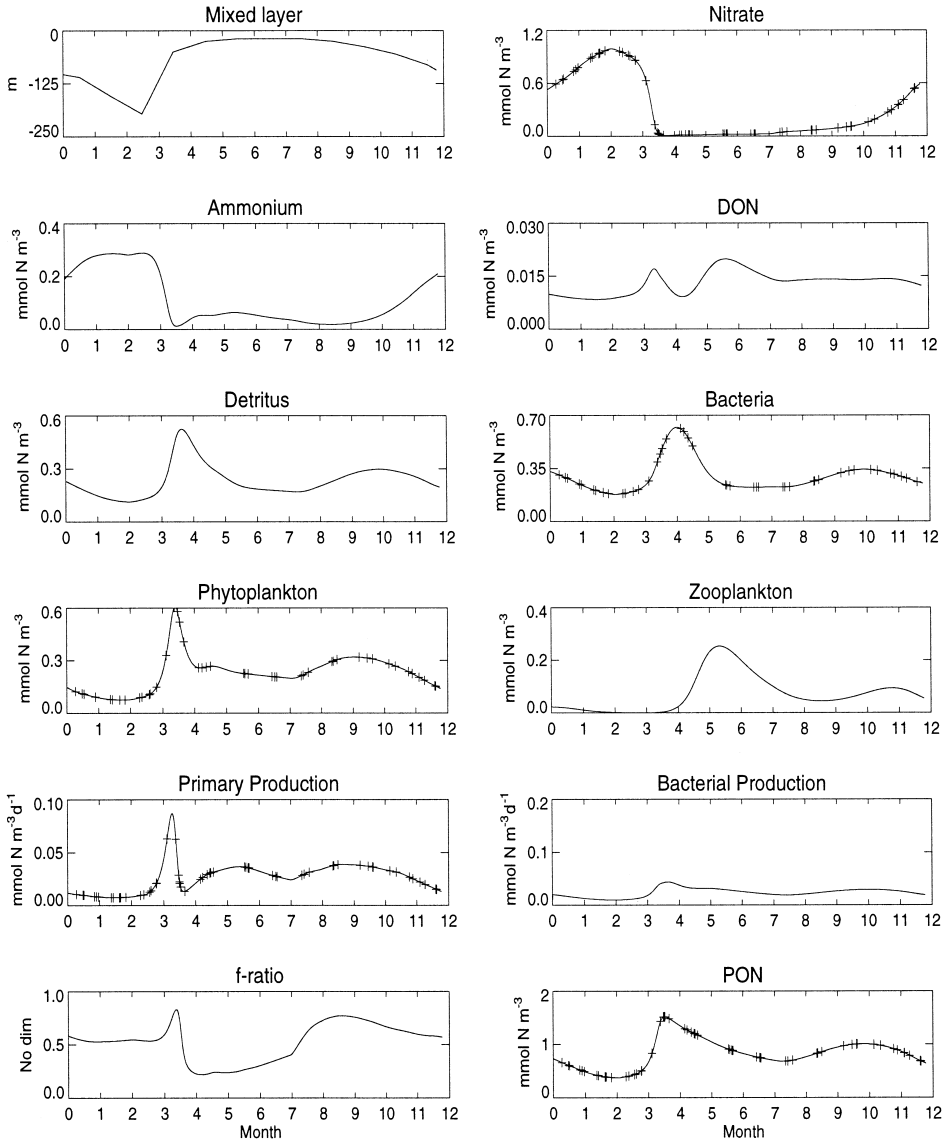

Fig. 2. Model results using Fasham et al. (1990) parameter values. The crosses represent the time of BATS observations used in the twin experiment assimilation process.

correlated and found a large uncertainty in the optimal parameters. However, his conclusions were based upon assimilation of observations from OWS Papa. There is no evidence that the inability to estimate individual parameters resides in their correlation, or is due to an inconsistency between the model and observations, or is due to the lack of information on some of the ecosystem components. Twin experiments using model-generated observations can address this dilemma for a given density of observations which mimics the frequency of BATS data collection. The twin experiment data set is guaranteed to be consistent with the model, free of measurement error, and expressed in the same units as the model results.

The twin experiment was designed as follows. The seasonally varying mixed-layer depth (MLD) was taken from climatological data for Bermuda to

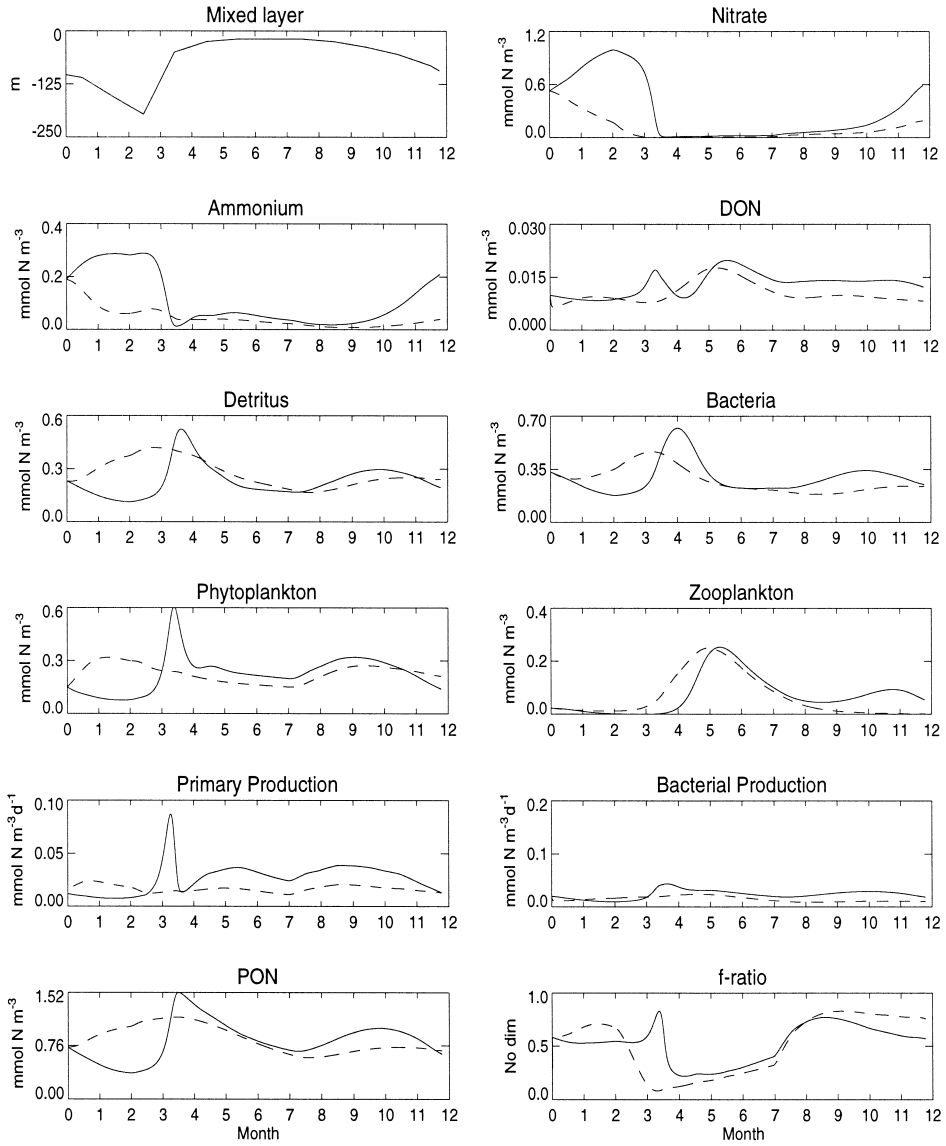

Fig. 3. Model results using Fasham et al. (1990) parameter values (solid line) and first-guess parameter values (dotted line).

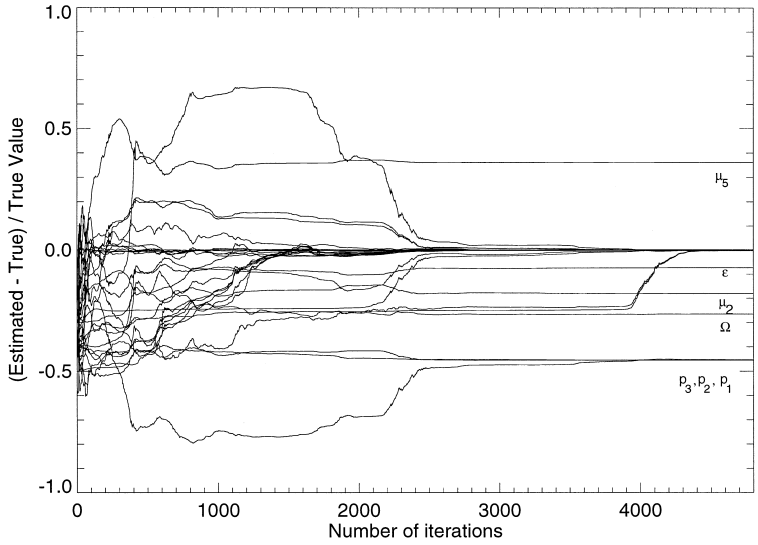

Fig. 4. Recovery of the Fasham et al. (1990) model parameters shown as the difference between the estimated model parameters and the true values normalized to the true value.

Table 3
Model parameters, true values and ratio of recovered to true value for the twin experiment

Parameter	Symbol	True value	Recovered/true value
Light attenuation due to water	k_w	0.04	1.0002794
Cross-thermocline mixing rate	m	0.1	1.0000571
Phytoplankton maximum growth rate	V_p	2.9	1.0005883
Initial slope of the P/I curve	α	0.025	1.0006963
Half-saturation for phytoplankton NO_3 uptake	K_1	0.5	1.0005164
Half-saturation for phytoplankton NH_4 uptake	K_2	0.5	1.0006672
Phytoplankton specific mortality rate	μ_1	0.09	1.000001
Light attenuation by chlorophyll-*a*	k_c	0.04	1.0003339
Phytoplankton exudation rate	γ_1	5%	0.99987815
NH_4 inhibition parameter	Ψ	1.5	1.0004517
Zooplankton maximum growth rate	g	1.0	1.0015351
Zooplankton assimilation efficiencies	β_1, β_2, β_3	75%	0.99940588, 0.99942592, 0.99868315
Zooplankton specific excretion rate	μ_2	0.1	0.81932421
Zooplankton specific mortality rate	μ_5	0.05	1.3610683
Zooplankton half-saturation for ingestion	K_3	1.0	1.0010011
Detritus fraction for zooplankton mortality	Ω	33%	0.73452456
NH_4 fraction of zooplankton excretion	ϵ	75%	0.92672895
Bacteria maximum growth rate	V_b	2.0	0.99822359
Bacteria specific excretion rate	μ_3	0.05	1.0000585
Detrital breakdown rate	μ_4	0.05	0.99999656
Bacteria half-saturation rate for uptake	K_4	0.5	0.99794159
NH_4/DON uptake ratio	η	0.6	1.0001300
Detritus sinking rate	V	1.0	1.0000254
Zooplankton food preference for phytoplankton	p_1	1.0	0.54542576
Zooplankton food preference for bacteria	p_2	1.0	0.54529131
Zooplankton food preference for detritus	p_3	1.0	0.54663190
Nitrate concentration below z = MDL	N_{bottom}	2.0	1.0000013

force the model with an observed annual cycle. The simulated observations were generated using the above ecosystem model with parameter values defined in Fasham et al. (1990) (see Table 2). Fasham et al. (1990) calibrated the model to reproduce the observations at Hydrostation S northwest of BATS. The model was run until it reached a steady annual cycle, which took 1 year of model time. The data were then taken from the following year of results. The simulated concentrations were subsampled to obtain observations at sampling times corresponding to the BATS observations from 1988 through 1992. Five years of sampling time were then folded into 1 year in order to increase the density of data since the time of data collection for any given month varies from year to year. The simulated observations are shown in Fig. 2. For the twin experiment, the parameter first guess was taken as 60% of the 'true' values (Table 2). The model was run for 1 year when it reached a steady annual cycle, and the data assimilation was done on the following year. This removes any effect that the initial conditions may have had on the recovery of the parameters. The second-year model results generated with the first-guess parameters are shown in Fig. 3. The main difference between the observations and the first-guess results resides in the absence of a well-defined spring bloom in the first-guess concentrations.

Assimilation of the simulated data resulted in the recovery of most of the 29 parameters (Fig. 4, and Table 3) except for the zooplankton-specific excretion and mortality rates (μ_2 and μ_5), the zooplankton food preferences (p_1, p_2, p_3), the fraction of the remineralized zooplankton grazing mortality (Ω), and the ammonium fraction of the zooplankton excretion (ϵ). Note that while those parameters were not recovered, each term in the equations containing those parameters (Eqs. (9), (19) and (20)) were correctly estimated. For instance, the sum of the zooplankton-specific excretion μ_2 and mortality μ_5 rates (Eq. (9)) was equal to 0.15 when using either the true values (Table 2) or the recovered values (Table 3). The same result occurred with the $\epsilon\mu_2 + (1-\Omega)\mu_5$ term in the ammonium equation, and the $(1-\epsilon)\mu_2$ term in the DON equation. This showed that the parameters involved in each of the terms mentioned are not independent and cannot be individually estimated.

The parameters which are involved in the model forcing terms were first recovered in order of relative importance of their contributions to the overall model solution. For example, the concentration of nitrate just below the mixed layer, N_{bottom}, which controls the injection rate of nutrients was first estimated. A signature of these recoveries is shown in the initial fast decrease of the value of the cost function and the

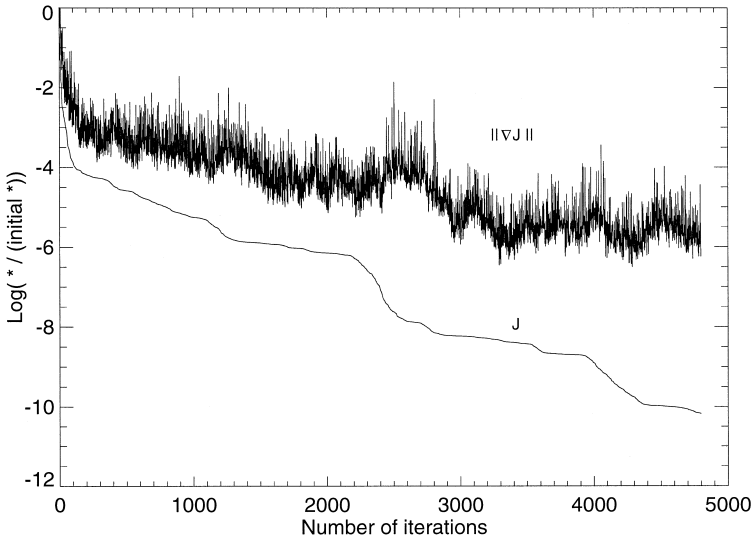

Fig. 5. Value of the cost function and the norm of its gradient as a function of the number of iterations using the Fasham et al. (1990) model.

norm of its gradient (Fig. 5). A second rapid decrease in the value of the cost function occurred around 2500 iterations. This corresponds to the recovery of various source and sink terms except for the ones controlling the bacterial source terms, a potentially important process within the microbial loop. The maximum bacterial uptake rate V_b and the half-saturation coefficient for the uptake K_4 are indeed the last parameters to be recovered. A similar recovery pattern was found, even in the best case, when the data set consisted of the model results for the seven ecosystem components at every time step (not shown). This suggests that the concentration observations contain only a weak signature of the processes related to the microbial loop. In other words, the recovery of the parameters associated with these processes is possible only after the model result errors due to the governing processes have been minimized.

After inspection of Eqs. (9), (19) and (20), the terms involving the non-recovered parameters were rewritten. The number of parameters was then reduced and assimilation of the simulated data led to the recovery of all the new parameters (Fig. 6). These changes did not alter the model dynamics, but did reduce the number of parameters which needed to be recovered. Specifically, $\mu_2 + \mu_5$ was replaced by μ_n, which parameterized the loss of zooplankton due to the combination of mortality and excretion. The term $\epsilon\mu_2 + (1-\Omega)\mu_5$ in the ammonium equation was replaced by $\delta_{NH_4}\mu_n$, which parameterized the flow of zooplankton loss into the ammonium pool. The term $(1-\epsilon)\mu_2$ was replaced by $\delta_{DON}\mu_n$, which parameterized the flow of zooplankton loss into the DON pool. The final simplified terms involved the zooplankton grazing terms, which were rewritten as:

$$G_1 = \frac{gZP^2}{K_3(P + p_1B + p_2D) + P^2 + p_1B^2 + p_2D^2} \quad (26)$$

$$G_2 = \frac{gZp_1B^2}{K_3(P + p_1B + p_2D) + P^2 + p_1B^2 + p_2D^2} \quad (27)$$

$$G_3 = \frac{gZp_2D^2}{K_3(P + p_1B + p_2D) + P^2 + p_1B^2 + p_2D^2} \quad (28)$$

where p_1, p_2 now denote the zooplankton preference for bacteria and detritus relative to the preference for phytoplankton.

Simplification of the model did not necessarily lead to a reduction in the number of iterations re-

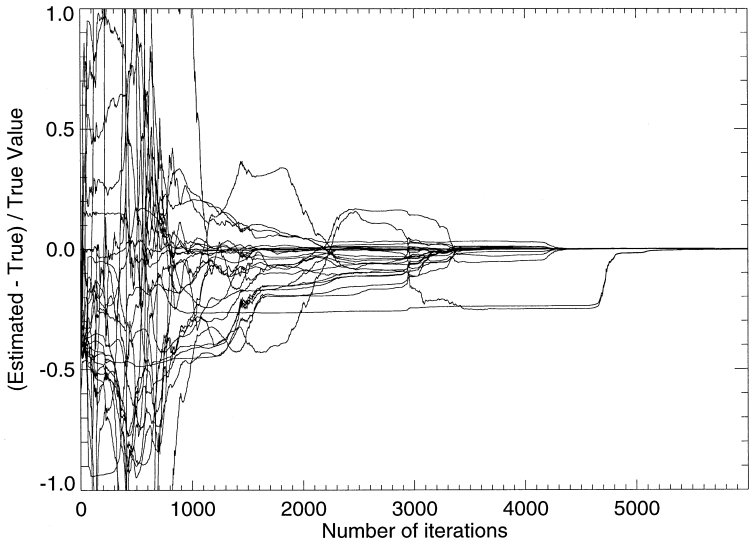

Fig. 6. Same as Fig. 4, except for the simplified model.

quired to recover the model parameters. While simplification of the model terms did not have an effect on the model results, it did have a non-negligible effect on the path of the recovery. Part of this was due to the differences in the first-guess model parameters and consequently in the first-guess model results. For example, the first-guess zooplankton concentration goes to zero in the modified model twin experiment. The other initial guess model results in the second experiment were also further apart from the modeled observations (Fig. 7) than that from the first experiment. Another reason for the difference in the two recovery experiments is that the path of the model recovery is sensitive to the first guess of the model. The change in the first-guess parameters caused a large variability in the parame-

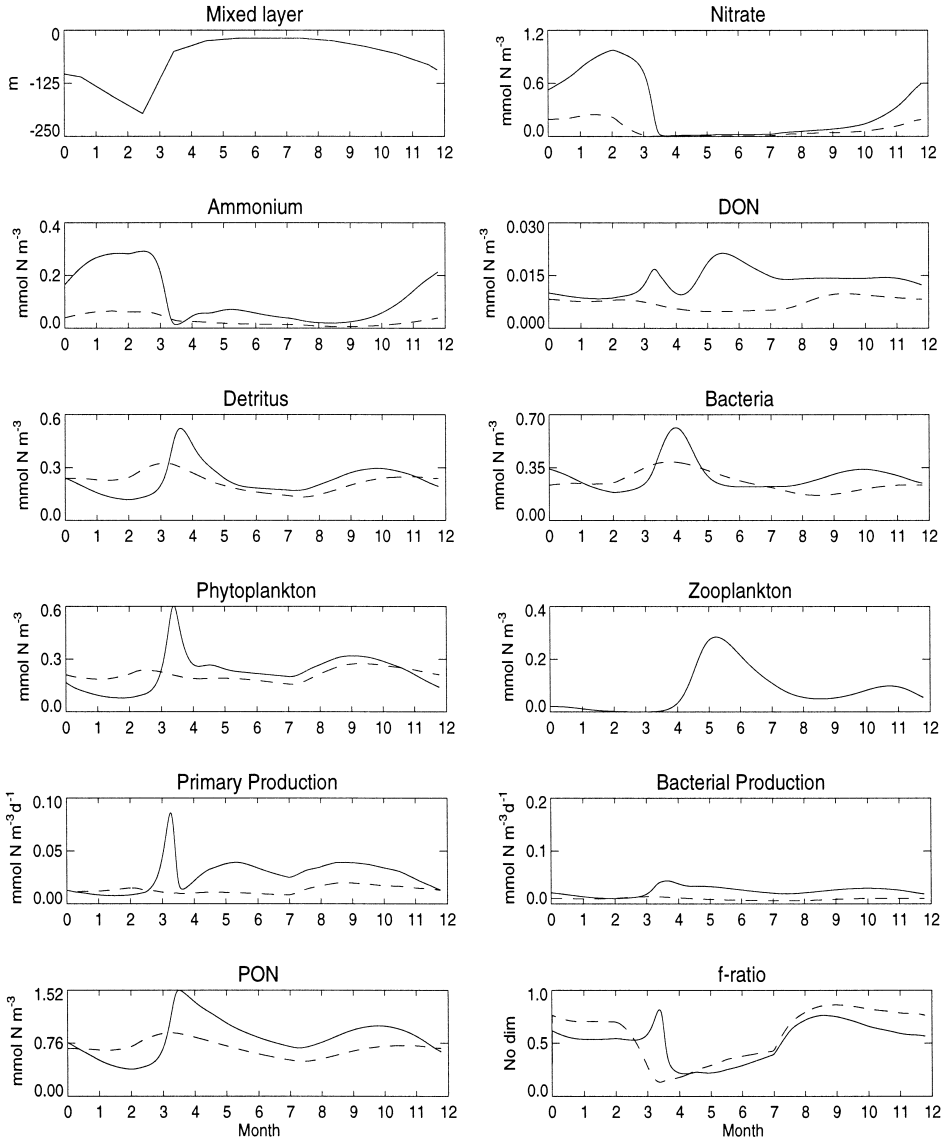

Fig. 7. Model results using Fasham et al. (1990) parameter values (solid line) and first-guess parameter values (dotted line) for the modified model twin experiment.

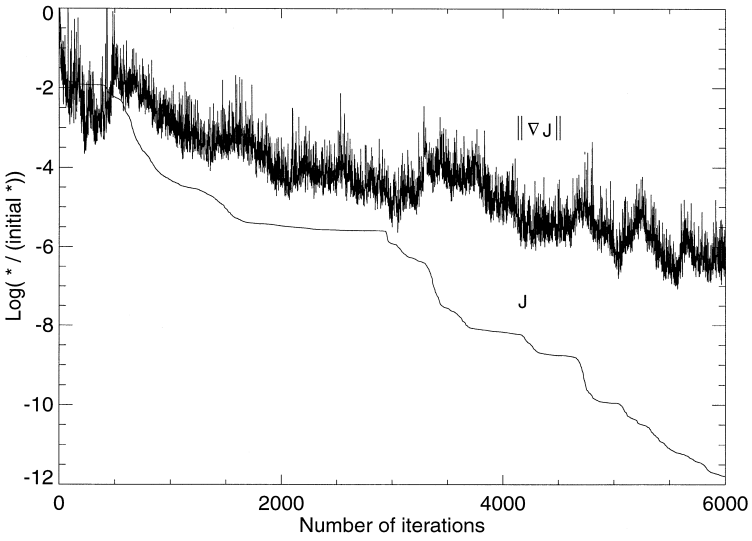

Fig. 8. Same as Fig. 5, except for the simplified model.

ter excursion during the first 1000 iterations of the assimilation process for the modified model twin experiment (Fig. 8). These differences between the two twin experiments also contributed to the higher number of iterations required for full recovery in the modified versus original model twin experiment. Note that while the number of iterations is high in both experiments, it could potentially be reduced by choosing different initial parameter guesses, by adding a penalty term in the cost function which takes into account a priori information on the parameters (Matear, 1995), or by assimilating either more or different types of observations (Tziperman et al., 1992; Lawson et al., 1996).

4. BATS data

In the previous section, we showed that all the parameters of the simplified model can be estimated using the same frequency and type of data as the observations from BATS between 1988 and 1993. In principle, the annual cycle of the ecosystem should then be modeled by estimating the model parameters using the actual BATS observations. The BATS data were obtained by vertically integrating the depth-dependent observations from the surface down to the prescribed MLD.

The first guess for the parameters was taken from the coefficients used in Fasham et al. (1990). The BATS data and first-guess model results are shown in Fig. 9. While the simulated nitrate time series compares well with the BATS data, all other model constituents compare poorly. For example, the simulated phytoplankton bloom occurs in late April and has a much larger amplitude than that observed in the BATS data. Also, the simulated nitrate increases faster in winter and spring than the observed one, and there is a peak in modeled bacteria, but no significant variation in the observed bacteria concentrations.

An attempt was made to estimate a realistic set of model parameters which could lead to a better fit to the actual data from BATS. However, the data assimilation failed to fit the model results to the observations. Consequently, the value of the cost function (not shown) could not be significantly reduced from that obtained from the first guess. A comparison between the model results generated with the estimated parameters and the BATS data is shown in Fig. 10. Similar results were obtained by Fasham and Evans (1995) when they used the Fasham et al. (1990) model and a non-linear optimization technique to model the ecosystem at the JGOFS station at 47°N 20°W.

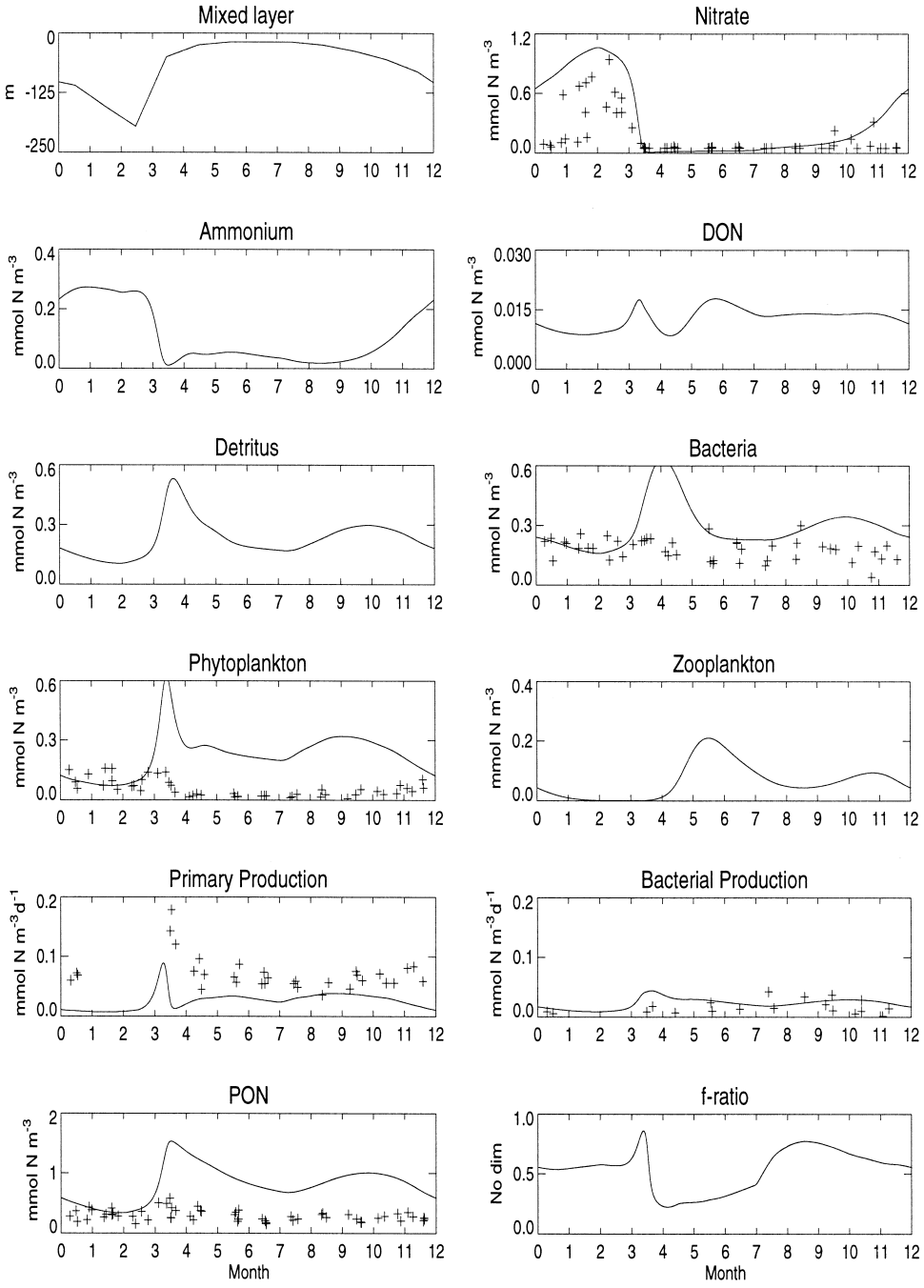

Fig. 9. BATS observations between 1988 and 1993 (crosses) and first guess model results (solid line).

Based upon the results from the twin experiments, we suspect that the ecosystem model does not adequately represent the ecosystem at BATS, and some of the model assumptions must be reconsidered. For instance, the primary forcing in the model is the rate of input of nitrate at the base of the mixed layer.

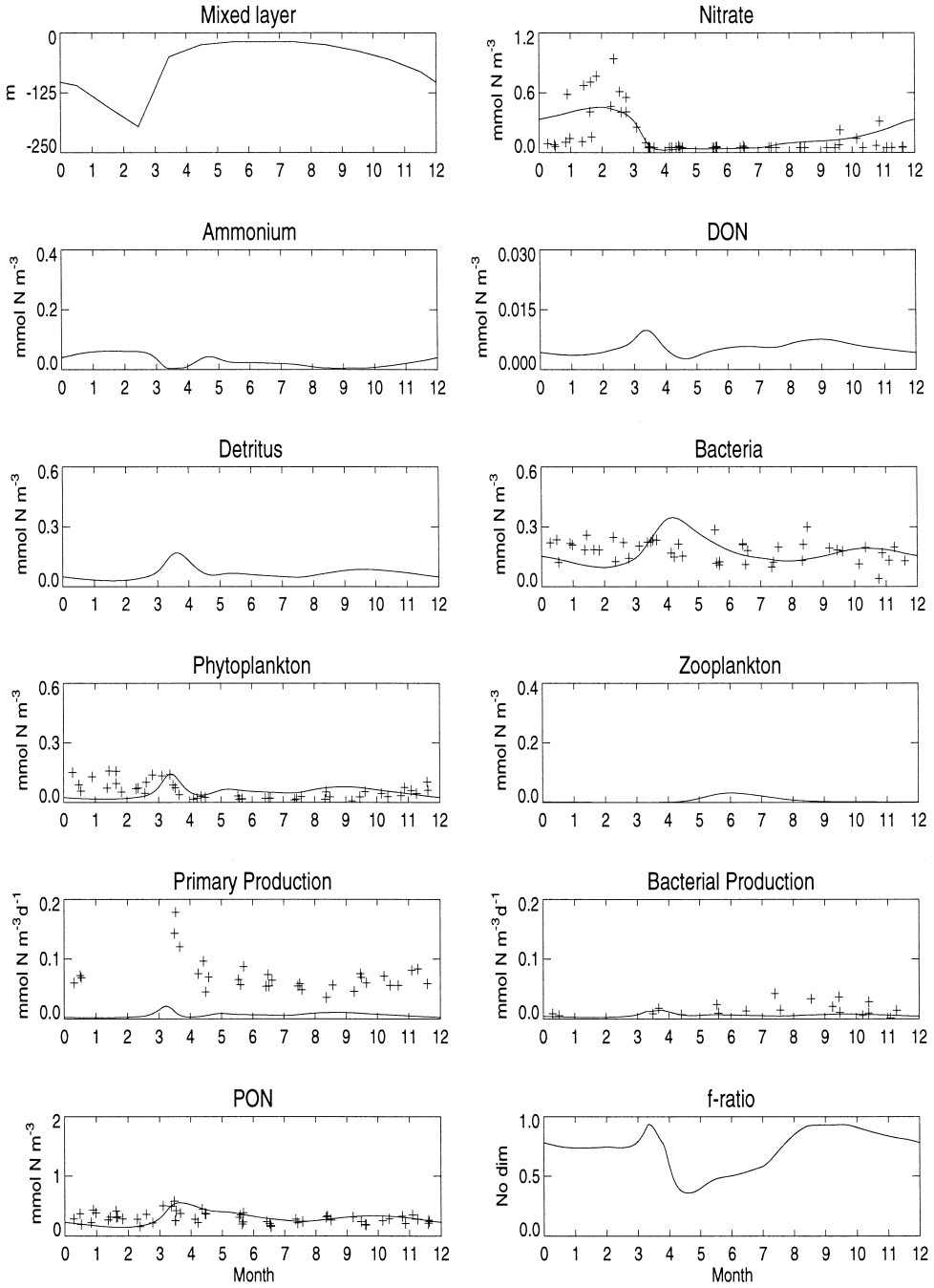

Fig. 10. BATS Observation between 1988 and 1993 (crosses) and results from data assimilation (solid line).

Fasham et al. (1990) assumed that the nitrate concentration below the mixed layer is constant over time and uniform with depth. The vertical profiles of nitrate observations at BATS show clearly that the nitrate concentration is approximately constant to 100 m deep and then varies linearly with depth.

Hurtt and Armstrong (1996) reformulated the nitrate input at the base of the mixed layer to account for the observed increase in nitrate concentration with depth. Similar reformulation was adapted by Fasham and Evans (1995) when they estimated the model parameters by assimilating the observations from the 1989 JGOFS North Atlantic Bloom Experiment. Another assumption in the model is that the chlorophyll-*a* to nitrogen ratio in the phytoplankton is assumed constant with time. This ratio is used to convert the phytoplankton nitrogen biomass into the measured chlorophyll concentrations. Phytoplankton cells can alter their chlorophyll-*a* concentrations so as to maximize the amount of energy absorbed utilized for photosynthesis. As a result, during periods of low light, such as winter, chlorophyll-*a* to nitrogen ratios are high, and during periods of high light, such as summer, chlorophyll-*a* to nitrogen ratios are low. This seasonal variability in the chlorophyll-*a* to nitrogen ratio was used by Hurtt and Armstrong (1996) and was shown to improve the Fasham et al. (1990) model with respect to BATS data.

Another source of the failure in the data assimilation can come from the assumptions made in converting the model simulated data into observations. For example, PON measurements at BATS measure the total particulate fraction of organic nitrogen in the water. However, much of this particulate material is known to be refractory and is not directly coupled to the seasonal cycle of the ecosystem. On the other hand, the model assumes that PON is composed entirely of phytoplankton, bacteria, zooplankton and detritus. Finally, the model was forced with a data set which contains interannual variabilities. The 5-year observations were indeed folded into 1 year to create the data set used in the assimilation process. This interannual variability might be a source of difficulty to estimate the annual cycle since the variability in the physical forcing, the mixed-layer depth, is not taken into account.

5. Summary

Based upon the results of the twin experiments, we can conclude that the type and frequency of the BATS observations are adequate to estimate the parameters in the Fasham et al. (1990) model and to simulate the annual cycle of the upper ocean ecosystem. The BATS observations were, however, not directly modeled, but were diagnostically obtained from the model results. We were also able show that some of the parameters were not independent. By addressing these dependencies, we simplified the model by reducing the number of parameters. In this case, all of the parameters were recovered, whereas dependencies between the model parameters prevented full parameter recovery.

These results demonstrate the utility of twin experiments to assess a priori the degree of model complexity that can be resolved with a given set of observations. Twin experiments are optimal since they are not confronted with measurement error and inconsistencies between the observations and the model. The addition of error to the observations can only hinder the rate of parameter recovery (Lawson et al., 1995). Additional observations may then be required to circumvent the problems caused by the measurement errors.

The results obtained from assimilating the actual BATS observations present clear evidence that the assumptions used in the Fasham et al. (1990) model need to be carefully reconsidered in order to model the pelagic ecosystem at BATS. In future work, we plan to modify the Fasham et al. (1990) model with respect to BATS data.

Acknowledgements

We thank the scientists and technicians from the U.S. JGOFS Bermuda Atlantic Time-Series program. In particular many thanks go to Drs. Anthony F. Michaels and Anthony H. Knap for their help and effort in compiling the BATS data set. We also thank J.C. Gilbert for supplying the code and documentation for the optimization subroutine, N1QN3. This research was supported by NASA EOS Contract NAGW-4596.

References

Armstrong, R.A., Sarmiento, J.L., Slater, R.D., 1995. Monitoring ocean productivity by assimilating satellite chlorophyll into ecosystem models. In: Powell, T.M., Steele, J.H., (Eds.),

Ecological Time Series, Chapmann and Hall, New York, pp. 371–390.

Carlson, C.A., Ducklow, H.W., Sleeter, Th.D., 1996. Stocks and dynamics of bacterioplankton in the northwestern Sargasso Sea. Deep Sea Res. II 43, 491–515.

Fasham, M.J.R., Evans, G.T., 1995. The use of optimisation techniques to model marine ecosystem dynamics at the JGOFS station at 47° N 20° W. Phil. Trans. Roy. Soc. Lond. B 348, 206–209.

Fasham, M.J.R., Ducklow, H.W., McKelvie, S.M., 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed-layer. J. Mar. Res. 48, 591–639.

Fasham, M.J.R., Sarmiento, J.L., Slater, R.D., Ducklow, H.W., Williams, R., 1993. Ecosystem behavior at Bermuda Station 'S' and Ocean Weather Station 'India': a general circulation model and observational analysis. Global Biogeochem. Cycles 7 (2), 379–415.

Ghil, M., Malanotte-Rizzoli, P., 1991. Data assimilation in meteorology and oceanography. Adv. Geophys. 33, 141–266.

Gilbert, J.-Ch., Lemaréchal, C., 1989. Some numerical experiments with variable-storage quasi-Newton algorithms. Math. Program. 45, 407–435.

Hurtt, G.C., Armstrong, R.A., 1996. A pelagic ecosystem model calibrated with BATS data. Deep Sea Res. II 43, 653–683.

Lawson, L.M., Spitz, Y.H., Hofmann, E.E., Long, R.B., 1995. A data assimilation technique applied to a predator–prey model. Bull. Math. Biol. 57, 593–617.

Lawson, L.M., Hofmann, E.E., Spitz, Y.H., 1996. Time series sampling and data assimilation in a simple marine ecosystem model. Deep Sea Res. II 43, 625–651.

Levitus, 1982. Climatological Atlas of the World. NOAA Prof. Pap. 13.

Malanotte-Rizzoli, P., 1996. Modern Approaches to Data Assimilation in Ocean Modeling. Elsevier, New York, 468 pp.

Matear, R.J., 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. J. Mar. Res. 53, 571–607.

Michaels, A.F., Knap, A.H., 1996. Overview of the U.S. JGOFS Bermuda Atlantic Time-series Study and the Hydrostation S program. Deep Sea Res. II 43, 157–198.

McGillicuddy, D.J. Jr., McCarthy, J.J., Robinson, A.R., 1995a. Coupled physical and biological modelling of the spring bloom in the North Atlantic (I): model formulation and one dimensional bloom processes. Deep Sea Res. I 42 (8), 1313–1357.

McGillicuddy, D.J. Jr., Robinson, A.R., McCarthy, J.J., 1995b. Coupled physical and biological modelling of the spring bloom in the North Atlantic (II): three dimensional bloom and post-bloom processes. Deep Sea Res. I 42 (8), 1359–1398.

Moisan, J.R., Hofmann, E.E., Haidvogel, D.B., 1996. Modeling nutrient and plankton processes in the California coastal transition zone. 2. A three-dimensional physical–bio-optical model. J. Geophys. Res. 101, 22677–22691.

Oguz, T., Ducklow, H., Malanotte-Rizzoli, P., Tugrul, S., Nezlin, N.P., Unluata, U., 1996. Simulation of annual plankton productivity cycle in the Black Sea by a one-dimensional physical–biological model. J. Geophys. Res. 101 (C7), 16585–16599.

Prunet, P., Minster, J.-F., Ruiz-Pino, D., Dadou, I., 1996a. Assimilation of surface data in one-dimensional physical–biogeochemical model of the surface ocean: 1. Method and preliminary results. Global Biogeochem. Cycles 10 (1), 111–138.

Prunet, P., Minster, J.-F., Echevin, V., Dadou, I., 1996b. Assimilation of surface data in one-dimensional physical–biogeochemical model of the surface ocean: 2. Adjusting a simple trophic model to chlorophyll, temperature, nitrate, and pCO_2 data. Global Biogeochem. Cycles 10 (1), 139–158.

Sarmiento, J.L., Slater, R.D., Fasham, M.J.R., Ducklow, H.W., Toggweiler, J.R., Evans, G.T., 1993. A seasonal three-dimensional ecosystem model of nitrogen cycling in the north Atlantic euphotic zone. Global Biogeochem. Cycles 7 (2), 417–450.

Sieracki, M.E., Haugen, E.M., Cucci, T.L., 1995. Overestimation of heterotrophic bacteria in the Sargasso Sea: direct evidence by flow and imaging cytometry. Deep Sea Res. I 42, 1399–1409.

Smedstad, O.M., O'Brien, J.J., 1991. Variational data assimilation and parameter estimation in an equatorial Pacific ocean model. Prog. Oceanog. 26, 179–241.

Spitz, Y.H., 1995. A feasibility study of dynamical assimilation of tide gauge data in the Chesapeake Bay. Ph.D. dissertation, Old Dominion University, Norfolk, VA, 167 pp.

Tziperman, E., Thacker, W.C., Long, R.B., Hwang, S.-M., 1992. Oceanic data analysis using a general circulation model. Part I. Simulations. J. Phys. Oceanogr. 22, 1434–1457.

Walsh, J.J., Dugdale, R.C., 1972. Nutrient submodels and simulation models of phytoplankton production in the sea. In: Kramer J., Allen, H., (Eds.), Nutrients in Natural Waters, Wiley, New York, pp. 171–191.